Digital Phenotyping and Cardiovascular Health

NCT04574882 (IRB #: 833699)

December 7, 2021

Protocol Details

Resubmission*
Yes

Hospital Sites
Will any research activities and/or services be conducted at a Penn Medicine affiliated hospital site?
Yes

Active Hospital Sites

Hospital of the University of Pennsylvania Pavilion (HUP) - East
Penn Presbyterian Medical Center (PPMC)
Hospital of the University of Pennsylvania (HUP) ***Primary***
Chester County Hospital
Pennsylvania Hospital (PAH)

Study Personnel

Principal Investigator

Name:	**MERCHANT, RAINA M**
Dept / School / Div:	**4302 - EG-Emergency Medicine**
Campus Address Mail Code	**4283**
Address:	**HUP** **3400 SPRUCE ST**
City State Zip:	**PHILADELPHIA PA 19104-4283**
Phone:	**215-615-3408**
Fax:	**215-573-2742**
Pager:	
Email:	**raina.merchant@pennmedicine.upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**11/27/2021**
Name of course completed :	**Good Clinical Practice: An Introduction to ICH (GCP) Guidelines**

Study Contacts

Name:	**SOUTHWICK, LAUREN E**
Dept / School / Div:	**4302 - EG-Emergency Medicine**
Campus Address Mail Code	**4283**
Address:	**HUP** **3400 SPRUCE ST**
City State Zip:	**PHILADELPHIA PA 19104-4283**
Phone:	**215-349-8382**
Fax:	**4283**
Pager:	
Email:	**Lauren.Southwick@uphs.upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**08/11/2024**
Name of course completed :	**Good Clinical Practice: An Introduction to ICH (GCP) Guidelines**

Name:	**TORRENTE, DENA M**
Dept / School / Div:	**4302 - EG-Emergency Medicine**
Campus Address Mail Code	**4283**
Address:	**HUP** **3400 SPRUCE ST**
City State Zip:	**PHILADELPHIA PA 19104-4283**
Phone:	**215-898-0339**
Fax:	**-**
Pager:	
Email:	**Dena.Torrente@uphs.upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**09/29/2024**
Name of course completed :	**Good Clinical Practice: An Introduction to ICH (GCP) Guidelines**

Name:	**MITRA, NANDITA**
Dept / School / Div:	**4635 - BE-Biostatistics Division**
Campus Address Mail Code	**6021**
Address:	**Department of Biostatistics University of Pennsylvania 212 Blockley Hall**
City State Zip:	**Philadelphia PA 19104-4865**
Phone:	**215-573-4467**
Fax:	**-**
Pager:	
Email:	**nanditam@pennmedicine.upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed :	

Other Investigator

Name:	**ASCH, DAVID A**
Dept / School / Div:	**4002 - SM-MG-Holding**
Campus Address Mail Code	**6218**
Address:	**423 Guardian Drive 1223 Blockley Hall**
City State Zip:	**PHILADELPHIA PA 19104-6021**
Phone:	**215-898-0102**
Fax:	**-**
Pager:	
Email:	**ASCH@WHARTON.UPENN.EDU**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed :	**CME Credit for POR Expedited Review - SOM**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed :	

Responsible Org (Department/School/Division):

4302 - EG-Emergency Medicine

Key Study Personnel

Name:	**MURGULESCU, VERONICA**
Department/School/Division:	**Health System**
HS Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**MENG, SABRINA**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**12/26/2027**
Name of course completed:	**CITI Good Clinical Practice (GCP) - OCR**

Name:	**SEHGAL, NEIL K**
Department/School/Division:	**School of Engineering**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**WEISSMAN, GARY E**
Department/School/Division:	**DM-Pulmonary, Allergy and Critical Care**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**SELTZER, EMILY**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**GROENEVELD, PETER W**
Department/School/Division:	**DM-General Internal Medicine**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**11/07/2024**
Name of course completed:	**Good Clinical Practice: An Introduction to ICH (GCP) Guidelines**

Name:	**FRANCISCO, ASHLEY**
Department/School/Division:	**PS-Neuropsychiatry**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**KLINGER, ELISSA**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**MANCHENO, CHRISTINA**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**04/15/2024**
Name of course completed:	**CITI Good Clinical Practice (GCP) - OCR**

Name:	**GUNTUKU, SHARATH CHANDRA**
Department/School/Division:	**Computer and Information Science**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**KAPADIA, ARYAN RAMCHANDRA**
Department/School/Division:	**Research Services**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**PELULLO, ARTHUR**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**KIMMEL, STEPHEN E**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**POR Recertification Quiz - Full Board Review - SOM**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**UNGAR, LYLE H**
Department/School/Division:	**Computer and Information Science**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**YANG, LINLIN**
Department/School/Division:	**SM-DN-Biomedical Graduate Studies**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**05/16/2025**
Name of course completed:	**CITI Good Clinical Practice (GCP) - OCR**

Name:	**SNIDER, CHRISTOPHER**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**JOSHI, SUKANYA M**
Department/School/Division:	**Research Services**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	

Name:	**SCHROEDER, DEVON**
Department/School/Division:	**OB-OBGYN RESEARCH-ADMIN**
HS Training Completed:	**No**
Training Expiration Date:	
Name of course completed:	
GCP Training Completed:	**Yes**
Training Expiration Date:	**10/03/2026**
Name of course completed:	**CITI Good Clinical Practice (GCP) - OCR**

Name:	**DISCHER, ADRIANA**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**
GCP Training Completed:	**Yes**
Training Expiration Date:	**12/17/2024**
Name of course completed:	**Penn CR: Full Onboarding: Good Clinical Practice: An Introduction to ICH GCP Guidelines (2HRS)**

Penn Intellectual Property*

To the best of the Principal Investigator's knowledge, does this protocol involve the testing, development or evaluation of a drug, device, product, or other type of intellectual property (IP) that is owned by or assigned to the University of Pennsylvania? Please refer to the Patent and Tangible Research Property Policies and Procedures.

No

Certification

I have reviewed the *Financial Disclosure and Presumptively Prohibited Conflicts for Faculty Participating in Clinical Trials* and the *Financial Disclosure Policy for Research and Sponsored Projects* with all persons who are responsible for the design, conduct, or reporting of this research; and all required Disclosures have been attached to this application.

Yes

Social and Biological Sciences

Study Instruments
Discuss the particulars of the research instruments, questionnaires and other evaluation instruments in detail. Provide validation documentation and or procedures to be used to validate instruments. For well know and generally accepted test instruments the detail here can be brief. More detail may be required for a novel or new instrument. For ethnographic studies identify any study instruments to be used (i.e. for deception studies) and describe in detail where, when and how the study will be conducted and who or what are the subjects of study. Note: For more information on how to conduct ethical and valid ethnographic research, follow the link For oral histories or interviews provide the general framework for questioning and means of data collection. If interviews or groups settings are to be audio taped or video taped describe in detail the conditions under which it will take place. Include a copy of any novel or new test instruments with the IRB submission.
After consenting, participants will complete a survey on a research laptop or on their own laptop at home (ThinkPad or MacBook air) computer (e.g. NHANES: SES [income, education], diet behavior and nutrition, physical activity and physical fitness, smoking and tobacco use; digital data use [Pew questionnaire]). Survey questions and citations are attached. For in-person enrollment, the survey will be collected on an incognito browser, https:// donate.centerfordigitalhealth.upenn.edu. The data collection platform was reviewed by Penn Medicine IS team. After they complete the survey, they will be asked to donate their social media and digital data on the ThinkPad laptop computer or their personal laptop computer using the same browser and url. All data extraction will be consistent with the terms of service of the digital platform. Any social media posts collected on Facebook, Instagram or other social media platforms will not be linked to the participant's name. A unique code links the participant's survey responses to the de-identified social media posts. All data will be kept in a protected field in a password-protected database. All participants will receive a unique ID to track survey responses. Participants will receive a copy of consent-related documents as well as detailed information about what data are being extracted, how those data will be used and shared among the research team and instructions on how to permanently delete any study apps that provide ongoing access. The study team will not communicate with participants via social media or publicly post that the participant is involved in the study. The study team will not have the ability to post any messages or content on behalf of the user. We received clearance from the Office of Clinical Research to not collect participants' Social Security Numbers, since compensation is $50. After the patient has completed enrollment, we will send them a follow-up summary about the status of the study (ie. how many participants are enrolled and the number of different data sources shared. After participants completed enrollment, we will re-contact participants to complete a short survey as related to objective 1. The study instrument will include questions related to health status, social media phenotypes, and AI perceptions, and patient activation. In addition to the above cohort, we offer the opportunity for survey respondents to enroll in a short interview (20-30 minutes) that will evaluate perceptions about use and engagement phenotypes, details about individual use of social media platforms, concerns and ways to mitigate concerns about phenotyping, potential uses of phenotypes, opportunities for patient centered design and implementation. Interviews will occur over 20-30 min in subgroups of 15-20 patients. Patients will consent to participate, and participation will be anonymous. Interviews will occur via virtual conference call and audio-recorded.

Group Modifications
Describe necessary changes that will or have been made to the study instruments for different groups.
No group modifications necessary.

Method for Assigning Subjects to Groups
Describe how subjects will be randomized to groups.
No randomization will occur.

Administration of Surveys and/or Process
Describe the approximate time and frequency for administering surveys and/or evaluations. For surveys, questionnaires and evaluations presented to groups and in settings such as high schools, focus group

sessions or community treatment centers explain how the process will be administered and who will oversee the process. For instance, discuss the potential issues of having teachers and other school personnel administer instruments to minors who are students especially if the content is sensitive in nature. Describe the procedure for audio and videotaping individual interviews and/or focus groups and the storage of the tapes. For instance, if audio tape recording is to be used in a classroom setting, describe how this will be managed if individuals in the class are not participating in the study. Explain if the research involves the review of records (including public databases or registries) with identifiable private information. If so, describe the type of information gathered from the records and if identifiers will be collected and retained with the data after it is retrieved. Describe the kinds of identifiers to be obtained, (i.e. names, social security numbers) and how long the identifiers will be retained and justification for use.

Participation should take 15-60 minutes during one session, but could take up to a week or more depending on the completion of all enrollment steps by the participant. Enrollment will take longer for participants opting to participate in the additional parts of the study (ie. additional questionnaire and/or focus group). Subjects may participate by having a research assistant approach them in person, via email, or over the phone. For in-person recruitment, participants will be approached during an in-patient visit stay following admission from the emergency department at HUP, HU pavilion, PAH, or PPMC or at outpatient settings like Penn Medicine Primary care clinics or Cardiology clinics in Philadelphia. Additionally, we will be reaching out to patients presenting to a Penn Medicine Chester County location, such as Lancaster General and Chester County hospital, remotely. Regardless of recruitment site, in-person recruitment, or remote recruitment, all patients will read and e-sign an online consent form, answer survey questions, and share their digital data. The survey and data donation platform lives within the existing structure of the Penn WayToHealth (WTH) platform, (https://donate.centerfordigitalhealth.upenn.edu/capture/) created by Co-Investigator Dr. Asch. Additionally, survey information will be captured through REDcap as well. WTH allows for secure patient consent and subject interaction.. All this data will be stored on our HIPAA secure shared drive and patient tracking information will be stored in Penn Box HIPAA compliant feature as well as REDcap. After the patient has completed enrollment, we will send them a follow-up summary about the status of the study (ie. how many participants are enrolled and the number of different data sources shared. For the Objective 1 subgroup participants, they will complete a short survey and have the option to opt-into a brief recorded Zoom interview on digital data norms and beliefs. Interviews will occur over 20-30 min in subgroups of 15-20 patients. Patients will consent to participate, and participation will be anonymous.

Data Management

Describe how and who manages confidential data, including how and where it will be stored and analyzed. For instance, describe if paper or electronic report forms will be used, how corrections to the report form will be made, how data will be entered into any database, and the person(s) responsible for creating and maintaining the research database. Describe the use of pseudonyms, code numbers and how listing of such identifiers will be kept separate from the research data.

All study data will be stored on the Health Service Research Data Center (HSRDC). As per the HSRDC Data Management Policy and Procedures Manual, password protection is used at the server and web portal levels for all transactions that allow entry and editing of data, or provide access to sensitive subject data or administrative privileges. Passwords will be managed to require all users to change their password within 90 days and strict rules will be implemented to require strong passwords. Additionally, all PHI data will be encrypted and linked to survey data through a study ID number, with linkage and de-encryption keys activated only by a user password for which a member of the research team has been given permission to access these sensitive data (Principal Investigator Dr. Merchant and project staff who are under the direct supervision of the PI). As per the HSRDC Data Management Policy and Procedures Manual, the database management on these servers is built with multiple layers of security and follows best practices for securing sensitive data. The main levels of security are fourfold and include machine physical security at the IT facility at which the servers are housed, physical server security in the highly restricted server containment room, electronic server security (firewalls, passwords, encryption) restricting access to the machine, and directory access controls restricting access to these particular data.

Radiation Exposure*

Are research subjects receiving any radiation exposure solely because they are enrolled in this protocol? (e.g. X-rays, CT, Fluoroscopy, DEXA, pQCT, FDG, Tc-99m, etc.)?IF YES, the protocol must be approved by the RRSC (Radiation Research Safety Committee). Consult EHRS web site:

www.ehrs.upenn.edu/protocols/radiohuman.html for more information.If you have questions, email jjesik@ehrs.upenn.edu or kavyap@upenn.eduIf your protocol includes Nuclear Medicine Procedures, the protocol must be reviewed by the Nuclear Med Operations Committee: https://redcap.link/NMOPS
No

Human Source Material*
Does this research include collection or use of human source material (i.e., human blood, blood products, tissues or body fluids)? IF YES, consult the EHRS web site: www.ehrs.upenn.edu/programs/bio/bbpathogens.html for information on OSHA Bloodborne Pathogens requirements (training, vaccination, work practices and Exposure Control Plan). If you have questions, call 215-898-4453.
No

Image Guided Biopsies*
Does the research involve imaging guided biopsy? IF YES, please contact the Clinical Imaging Core. See https://www.med.upenn.edu/cbi for more details. Any questions should be directed to the Director of Research Operations, Dept of Radiology, Kathleen Thomas.
No

Computerized Tomography (CT) Studies*
Does the protocol involve CT scans that are not considered standard of care and are being performed for research purposes?IF YES, complete the CACTIS Committee Application: https://is.gd/CACTIS and consult CACTIS website: http://www.uphs.upenn.edu/radiology/research/labs/cactis/ for application requirements.
No

CAMRIS and MRI Studies*
Is an MRI scan being performed for research only and NOT considered standard of care (example: specific scanner, parameters or solely for the purposes of research)?NOTE: Research/non-standard use of MRI may include but is not limited to any of the following: Situations in which MRI results may impact subjects current clinical care plan or treatment decisions, such as:The study requires a customized report with specifics regarding the study protocol (i.e., specific measurements or details); Introduction of a device of any kind during the MRI that is not used during a 'standard of care' type scan.Your MRI is not consistent with standard care time points for MRI imaging.Your MRI is not paid for by insurance.IF YES, consult CAMRIS website: https://www.med.upenn.edu/camris/application-and-faq.html for application requirements and required institutional consent form language.
No

Cancer Related research not being conducted by an NCI cooperative group*
Does this protocol involve cancer-related studies in any of the following categories? Therapeutic, Prevention, Supportive Care, Screening, Early Detection, or Diagnostic, Epidemiologic, Observational, Outcome, Ancillary or Correlative. For a description of these categories, see http://www.ctsrmc.org/submitting_a_protocol.php NCI Cooperative Groups are as follows:Alliance for Clinical Trials in OncologyNCI Clinical Trials Group (Canadian Cancer Society) (NCCTG)Children's Oncology Group (COG)NRG Oncology GroupECOG-ACRIN Cancer Research GroupSouthwest Oncology Group (SWOG) IF YES, the protocol must be submitted to the Cancer Center's Clinical Trials Scientific Review Committee for scientific review and approval prior to obtaining IRB approval. Consult the CTSRMC website: www.ctsrmc.org for application requirements
No

HIPAA / Protected Health Information
Does the research proposal involve accessing (viewing / using), collecting, or disclosing of protected health information (PHI) directly from participants or their medical or dental record for research purposes?
Yes

CHPS Resources*
Does the research involve CHPS resources?
No

HUP Inpatient Nursing Resources
Does this research include an inpatient admission at HUP?
No

If the answer is YES, indicate which items is is provided with this submission:

Modified research informed consent document that incorporates HIPAA requirements

Use of UPHS services*
Does your study require the use of University of Pennsylvania Health System (UPHS) services, tests or procedures , whether considered routine care or strictly for research purposes? (UPHS includes all Penn hospitals and clinical practices, including the Clinical Care Associates network of community practices). Examples of UPHS services/tests/procedures includes the Clinical Translational Research Center (CTRC), laboratory tests, use of the pathology lab, cardiovascular imaging tests or radiology imaging tests (whether being billed via the Service Center or through UPHS), other diagnostic tests & procedures and associated professional services, etc.
No

Veteran's Affairs (VA) Patients or Subjects
Does your study involve data from Veteran's Affairs (VA) patients or subjects?
No

If yes, was this approved by the Philadelphia VA?

No

Out of State Research
Will any Penn personnel conduct any research activities outside of the State of Pennsylvania?
No

Research involving Virtua Health
Will any Penn personnel conduct any research activities at a Virtua Health site location, OR in collaboration with Virtua Health System personnel, OR using any Virtua Health System resources (e.g., medical records)?
No

Primary Focus*
Research on human data sets (e.g. medical records, clinical registries, existing research data sets, medical administrative data, etc.)

Protocol Interventions

x	**Sociobehavioral (i.e. cognitive or behavioral therapy)**
	Drug
	Device - therapeutic
	Device - diagnostic (assessing a device for sensitivity or specificity in disease diagnosis)
	Surgical
	Diagnostic test/procedure (research-related diagnostic test or procedure)
	Obtaining human tissue for basic research or biospecimen bank
x	**Survey instrument**
	None of the above

The following documents are currently attached to this item:

There are no documents attached for this item.

Sponsors

Business Administrator

Name:	**RIZIO, SUZANNE**
Dept / School / Div:	**4302 - EG-Emergency Medicine**
Phone:	**215-898-9108**
Fax:	**215-898-0868**
Pager:	
Email:	**rizio@pennmedicine.upenn.edu**

Department budget code

400 - 400 - 4 - 577869 - 5316 - 2000 - 3520

Funding Sponsors

Name:	**NATIONAL INSTITUTES OF HEALTH**
Type:	**UPENN Federal**

Funding sponsors billing address
If you have selected a commercial or industry sponsor, please provide the appropriate address and contact information for the Sponsor for the purposes of billing for IRB review fees (initial review, continuing review and convened modification fees apply here). If the Sponsor is not industry/ commercial, this information is not necessary to provide with your application.
National Heart, Lung, and Blood Institute Building 31 31 Center Drive, Bethesda, MD 20814 United States

Funding sponsors gift
Is this research being funded by a philanthropic gift?
No

Regulatory Sponsor

IND/IDE Sponsor

none

400 - 400 - 4 - 577869 - 5316 - 2000 - 3520

Industry Sponsor

None

Project Funding*
Is this project funded by or associated with a grant or contract?
Pending

Sponsor Funding
Is this study funded by an industry sponsor?
No

Status of contract

The following documents are currently attached to this item:

There are no documents attached for this item.

Multi-Center Research

Penn as lead
1. Is this a multi-center study where Penn is serving as the Lead Site or the Penn PI is serving as the Lead Investigator?
No

Management of Information for Multi-Center Research

Penn irb of record
2. Is this a multi-center study where the Penn IRB will be asked to serve as the IRB of Record for other external study sites?
No

Other Sites

No other sites

Protocol

Abstract
Cardiovascular disease is the leading cause of death in the US. While prevention approaches have improved longevity of patients, risk factors and adverse health behaviors (e.g., physical inactivity, smoking) are highly prevalent. The logistics and practicalities of meeting the goal of ideal CV health have not been clearly elucidated. Practice guidelines recommend using the Framingham risk score (FRS) or other risk prediction tools to classify patients risk of CV disease. These models however are imprecise and there is increasing focus on identifying markers that provide better measures of risk. As digital platforms are increasingly used to document lifestyle and health behaviors, data from digital sources may provide a window into manifestations of novel risk factors and potentially a better characterization of existing risk factors.

Objectives

Overall objectives
This study has three main objectives. (1) Identify and characterize topics and features derived from digital data (e.g., social media, online search, mobile media) which are associated with CHD and related risk factors and develop phenotypes of how patients use and engage with social media and then determine the association of these phenotypes with health status, social media perceptions, and patient activation. (2) Quantify and validate the incremental benefit of adding variables from patients' digital data to conventionally-derived predictive models of CHD (e.g. Framingham risk score/American College of Cardiology ASCVD Risk Estimator Plus) using a retrospective cohort of patients with and without heart disease. (3) In a cohort of patients with and without heart disease, develop and evaluate models that use digital media data to predict health care utilization (e.g. cost). (4) Evaluate the impact of dynamic health events (e.g. COVID-19) on modifiable CV health behaviors (smoking, diet, activity) using population level-digital media data (e.g. Facebook, Twitter)

Primary outcome variable(s)
The primary outcome variable for the first objective will be topics and features (derived using the Latent Dirichlet Allocation [LDA] method for clustering language data) as well as a model-based approach of latent class analysis (LCA). We hypothesize: H1: There will be a difference in themes and topics posted by patients with heart disease and for patients without heart disease and use and engagement

phenotypes reflecting quantity, quality, and interaction of posts will be associated with health status, social media perceptions, and patient activation. The second objective outcome variable will be any CHD event. We hypothesize: H2: Adding digital data variables to conventionally-derived predictive models will improve the AUC for forecasting/predicting a CHD related event in a patient. The third objective outcome measurewill be health care utilization measured by cost. We hypothesize: H3: Unique digital data variables identified for patients with heart disease and without heart disease will better predict future health care utilization costs for the health system. The fourth objective outcome measure will be the impact of dynamic health events on CV health behaviors. We have the following three hypotheses: H4,1: Digital media data will provide valid measures of health behaviors when compared with traditional sources. H4,2: Variations in digital data about modifiable CV health behaviors will correlate with dynamic health events. H4,3: Modifiable CV health behaviors (smoking, poor diet, and reduced physical activity) will increase more in response to dynamic health events in low income and minority communities.

Secondary outcome variable(s)
Given the correlational nature of this research the distinction between primary and secondary outcome variables is not particularly apparent.

Background
Cardiovascular disease is the leading cause of death in the US. While secondary prevention approaches have improved longevity of patients, risk factors and adverse health behaviors (e.g., physical inactivity, smoking) are highly prevalent, and in most contemporary series, less than 1% of adults meet all factors of ideal CV health. The logistics and practicalities of meeting the goal of ideal CV health have not been clearly elucidated. Practice guidelines recommend using the Framingham risk score (FRS) or other risk prediction tools to classify patients' risk of CV disease. These models however are imprecise and there is increasing focus on identifying markers that provide better measures of risk. As digital platforms are increasingly used to document lifestyle and health behaviors, data from digital sources may provide a window into manifestations of novel risk factors and potentially a better characterization of existing risk factors. While it seems like a cliche to mention the profound impact of digital data on everyday lives, there is indeed great substance in the opportunities these new media provide for understanding behavioral, social, and environmental determinants of health. Health is shaped most by the behaviors, events, and circumstances that occur outside of traditional health care contact. Even a patient with a chronic illness may spend only a few hours a year in front of a doctor, but the same patient will spend over 5,000 waking hours a year doing everything else. And it is during those other 5,000 hours that so many of the determinants of people's health play out. Digital footprint data offer a window into what people do in those 5,000 hours. Social and digital information in the form of posts, photo likes, Internet searches, steps walked, and other online footprints can reveal a detailed narrative about an individuals day-to-day activities and behaviors. Google search queries have been used to measure trends in influenza, predict prevalence of disease, and estimate congestive heart failure admissions. Social media has been used to track infectious disease outbreaks, cardiac arrest geography, heart disease mortality rates, access to healthy food, and depression. Steps data from sensors (e.g. iPhone, Android accelerometers) have been used for real-time tracking and as a tool to improve physical activity in adults with cardiac disease. The widespread interest in precision medicine based largely on people's genomic identities took hold once we had a way to capture genomic information from DNA sequencing. Digital data may be complementary and provide new signal as: [1] social information is now more accessible, not requiring the bio samples that limit the reach and generalizability of genomic data; [2] digital platforms often come with their own 'return to sender' pathway that offers opportunities for intervention, making information derived from it more actionable. Digital platform users are self-selected and thus reflections of what people reveal organically can't represent the population as a whole, but it is clear that the digital divide is narrowing, and that utilization of digital platforms is growing exponentially. The fastest growing group of users of social media is over the age of 65. African-Americans, Latinos, and those in urban populations are in fact overrepresented on some social platforms (e.g., Twitter) relative to the general population. Google use is ubiquitous with more than 1.17 billion people performing more than 2 trillion searches per year. More than 50% of individuals in the US are iPhone users and 1 in 5 Americans owns a wearable device. Of central importance to this proposal is that individuals from diverse backgrounds are increasingly using these platforms for communication about health and healthcare. The speed with which data are generated creates potential for earlier signal about disease development or exacerbation. The regular production of data also allows for tracking of health in real-time or over expanded periods time. The variety of data offers promise as this information provides different insights than data from traditional surveys or claims. Another

opportunity is in the ability of digital platforms to not only provide markers of health but also serve as platforms that can be used for direct intervention. For example, Facebook now allows users to flag posts within their network that they think may suggest suicidal ideation. Facebook then anonymously provides resources for potential individuals at risk. To expand the scale of this type of intervention for mental health and other disorders would require calibrated algorithms with high sensitivity and specificity. The potential in CV health is in tracking, codifying, and better understanding the hard to measure lifestyle choices and exposures related to diet, exercise, smoking and others which can significantly contribute to development and progression of heart disease. At present measuring many of these behaviors is dependent on self-report and recall. Yet, posts or images from digital media could better inform a patient-provider discussion about how to change for example actual dietary choices and consumption. Standing in the way of this potential promise are a host of questions that require further scientific study about the social acceptability and logistic feasibility of collecting ongoing streams of data from individuals, about the analytic approaches essential for identifying associations between those data and more conventional representations of individual health, and about the application of those associations for research and for individual and population health. Further needed is a better understanding of the accuracy of the data, motivations for posting, representativeness (are posts actually by the person assigned), missing data (what behaviors are not captured intentionally or unintentionally) and others. The proposed project represents an entirely new area of study in looking at multiple digital media outputs and their association with CV risk and health care utilization. Understanding the collection, interpretation, and use of these new data streams represents a pioneering exploration of potentially great and enduring value. Ideal cardiovascular health (CVH) is based on 7 modifiable CV factors and behaviors (smoking, BMI, diet, physical activity, smoking status).74 Low risk with respect to these collective measures is associated with reduced CV and stroke incidence and mortality.87 Only 5% of US Adults have 6 metrics in the ideal range.73 Modifiable CV behaviors are tracked through different validated nationally representative surveys (e.g. NHANES) but the output from these surveys often lag significantly in their availability for researchers and providers. Therefore, there are significant limitations in being able to develop public health interventions to address population level changes in health behaviors in near real time. Less is known about how modifiable health behaviors change in response to dynamic public health events. On March 11, the World Health Organization declared novel coronavirus (COVID-19) a global pandemic and identified that it was associated with an infodemic of information online.88 As millions of Americans were required to shelter in place, health behaviors were notably altered and social media use exponentially increased.89 Quarantine and sheltering in place have been associated with increased stress, anxiety, exacerbation of mental health conditions, and alterations in diet, weight, and physical activity.90 In the era of COVID-19, disparities in care and outcomes have already been described.91, 92 The impact of COVID-19 is predicted to be longstanding due to alterations in the economy, education, and other areas of society which impact health. We will use advanced computational models for identifying and codifying health behaviors by region over time and in real time. This can inform strategies that health systems and public health departments can deploy for addressing the immediate and long-term CV health of the patients and communities they serve. Ideal CV health is ideal but making lifestyle and behavioral changes is hard.99, 100 Digital interventions can be used to enhance self-management of diabetes and modifiable behaviors like smoking cessation.101-103 Of increased focused, engagement phenotypes that help explain how individuals use and engage with technology can be used for tailoring behavioral interventions and sustained behavioral change.76 While marketers routinely use this information for increasing interest in products, there is limited data on how patient phenotypes (developed from consenting patients) could be used for personalized health interventions using social media data and social media platforms.

Study Design

Phase*
Not applicable

Design
We will identify patients ages 30-74 with and without CVD) within the last 5 years who receive care through the University of Pennsylvania Health System or receive care in another health system in the United States (U.S.). For patients with CVD, we will be using the following ICD codes for identification: ICD 10: I63, I20-I25. For patients without CVD, please find attached the list of ICD 10 codes we will use for identification. UPHS is the largest health system in Eastern Pennsylvania with a diverse patient population of more than 500,000 patients with CV disease. We will recruit both in and out-patients as study participants. For both the inpatient population and outpatient population within

UPHS, we will use Epic workbench and other Epic screening tools such as the Best Practice Alerts (BPA) to screen for eligible patients. BPAs can be configured to identify patients given a set of inclusion criteria, in real-time, and send a silent in-basket message to allocated study personnel. Users gain access to these alerts through an in-basket message pool within PennChart, BPA distribution and modification privileges can be controlled (and limited) based on security configurations. Penn Medicine sites identified in the BPA include PCAM, HUP, PPMC, PAH, CCH, Cardiology HVC Consultative Clinic, Cardiology HVC Clinic, PCAM, Cooper/PIMA Internal Med, Penn Cardiology Bucks County, Penn Cardiology South Philly, Penn Cardiology Valley Forge, Penn Medicine Bala Cynwyd, Penn Heart and Vascular Center Washington Square, Penn Cedar and Penn Heart and Vascular Center Radnor, PA. For patients who are identified outside of UPHS, we will solely rely on self-report diagnosis for identification. Potentially eligible patients will either be approached or emailed/called about the opportunity to participate in the study either in person during an in-patient visit (4,734,948/annually) or outpatient visit at Penn Medicine hospital, including the Emergency Rooms at the Hospital of the University of Pennsylvania (HUP), Penn Presbyterian Medical Center, and Pennsylvania Hospital, or at one of the Outpatient Internal Medicine Clinics in Philadelphia, or contacted remotely (over a secure calling phone application or emailed). Additionally, we will contact patients remotely at various Penn Medicine Chester County locations such as Lancaster General and Chester County hospital. For initial interest participation emails and follow-up emails, we will be using a shared Penn Medicine email account through our Center (dydhearthealth@pennmedicine.upenn.edu) to reach out to those folks. In addition to emailing and calling patients, we will be using My Penn Medicine (MPM) which is UPHS's secure patient portal to message identified patients about interest in our study. Our message will use the same template as our initial outreach email. Another form of remote outreach we will be using is Way to Health (W2H), which is a HIPAA secure two-way texting platform, where we will text message identified patients our initial outreach message about our study (reference the text message automated flow attached). If patients are interested, they will be texted our eligibility survey link to complete. We will also use this HIPAA secure text messaging platform to follow up with participants related to various study participation (ie. re-sharing of data through a secure link, scheduling confirmation). Interested participants who are eligible will be asked to fill out a short survey asking what days and times work best for a team member to give them a call to walk through what is involved in participating in the study. We will use the Academic Associate (AA) Program in the Emergency Department (ED) at the Hospital of the University of Pennsylvania, Presbyterian Medical Center to identify patients. The AA Program provides premedical and pre-health post-baccalaureate students an opportunity to engage in clinical and translational research activities. The AAs have Immunization records (measles, mumps, and rubella [MMR] and varicella [chickenpox]) and TB screening (PPD test or equivalent), and influenza (in season) and complete a web-based course sponsored by the University of Pennsylvania on the protection of human subjects in research. Using Epic tools and AAs, we will screen all patients aged 30-74 examining their chief complaint recorded in Epic. If patients present cardiovascular-related chief complaints, the patient will be noted in a screening log, and if admitted, note their in-patient ICD-10 admission code. Similarly, for patients aged 30-74 who have a non-cardiovascular-related chief complaint, the patient will be noted in a screening log and if admitted, note their in-patient ICD-10 admission code. For in-person recruitment and enrollment in the outpatient setting, outpatient Penn Medicine clinics (e.g. Internal Medicine, Primary Care, Cardiology), RAs will review the weekly appointment list to review if any patients meet either case and control requirements. RAs will be in communication regularly with the nursing coordinator, office manager, charge nurse, nurse/ physician team to determine the daily census and appropriateness of approaching patients. In both in and out-patient settings, the AAs or CDH research staff will approach and will introduce the study in brief and if the patient is interested in enrolling. They query eligible patients about their interest in study enrollment (i.e. sharing their social media, online search, or step tracking data and allowing access to their EMR data and health insurance data). For remote recruitment and enrollment, identified patients via the BPA and EPIC reporting workbench will either be emailed, called, sent an MPM message, or a W2H message in regards to interest in participating in this research study. Initial interest participation emails and all follow-up emails will be sent using our Centers shared Penn Medicine email account (dydhearthealth@pennmedicine.upenn.edu), MPM messaging, or W2H messaging. All recruitment and follow-up emails will be A/B tested, also known as split testing, the process of comparing two identical versions of an email. Potential variations include the order of information, email headers, and signatures. NOTE: In addition to the research team reaching out to potentially eligible patients, we will be using flyers, we will post our study on Penn Medicine Lancaster General Health's various social media sites (ie. Facebook, Instagram, Twitter, LinkedIn, etc.), and we will be posting our study on iConnect for interested patients to reach out to us regarding participation in the research study. We will be using iConnects pre-screener feature to assess patients for pre-eligibility for our study. These pre-

screener questions will be the same as the questions already built into our eligibility survey. If the patient is interested, they will receive a link to an eligibility survey to complete either over the phone with a team member or on their own time. The link will be sent via our secure shared Penn medicine email, MPM, or W2H on our HIPAA secure survey platform. If the patient is eligible and still interested, they will be asked to fill out a baseline survey at a time that works best for them. At that time, interested participants will be asked to complete the informed consent HIPAA authorization process at that time. The HIPAA authorization will also include detailed information about the request for permission to collect cost data. For example We are asking permission to collect information on your Medicare and/or insurance coverage and on your cost of care. This information will be collected directly from your insurance and medical record. Information will be collected for up to 5 years. If you agree to participate in this study the following information will be collected: type of health insurance, health insurance provider, health insurance policy, health insurance group number, and the policyholder's name and date of birth will be collected. If we aren't able to obtain your health insurance policy number, we will either call you to obtain your policy number or you can fill out a secure survey through REDcap that will allow you to input your policy number. For in-person enrollment, after the patient signs the informed consent the AA or CDH staff will complete the baseline survey and continue the enrollment process. Patients will complete several surveys on a research laptop (ThinkPad or MacBook air) and/or an iPad (e.g. NHANES: SES [income, education], diet behavior and nutrition, physical activity and physical fitness, smoking, and tobacco use; digital data use [Pew questionnaire]). For patients outside of UPHS, we will ask them to upload screenshots and/or self-report information from their office visit summaries, labs, vital signs, ED and inpatient hospital admission information since we will not have access to their EHR. For in-person enrollment, the survey will be collected on an incognito browser through either https:// donate.centerfordigitalhealth.upenn.edu or REDcap. Once the survey is completed, the patient will reach a donate data page on the same browser, https:// donate.centerfordigitalhealth.upenn.edu or be taking to another survey to share their digital data, https:// donate.centerfordigitalhealth.upenn.edu, on one of the research laptops. For remote enrollment, participants will receive a link via email from our HIPAA secure survey platform to complete the baseline survey either on their own time or over the phone with a study team member. Interested participants will be asked to fill out a brief survey, via Redcap, asking what days and times work best for them to set-up a phone call with one of our team members to go over what is involved in participating in the study. They can either complete the enrollment process on the phone with the team member or on their own time. Subjects who meet case or control criteria, complete the consent process, and opt to share their social media data with the research team, will be instructed on how to share this information by logging into apps for Facebook and Instagram that have been created solely for the purpose of research. They will first be asked to create an account on our secure digital data platform/ application (https:// donate.centerfordigitalhealth.upenn.edu). Once they have confirmed information in the data collection survey, they will be directed to a Facebook application called Penn Image Station that was designed for the purposes of this study. The Penn Image Station application will gather information from their Facebook public profile, status updates, personal description, and likes. The application will not affect their usual use of Facebook nor will it post on their page. Facebook data will be collected prospectively for 60 days. If the participant shares their Instagram posts, they will be directed to an Instagram application called Penn Social Mediome. They will read a brief statement, either agree or disagree with the statement, and then log on to their Instagram account. Their Instagram posts and photos will be extracted by the Penn Social Mediome application and will be collected for 60 days prospectively. If our data collection application is not working, there will be a back-up plan set in place. We will ask participants to share their Instagram usernames so we to pull data manually from their account. If there are technological difficulties with our data collection application, we will reach out to participants at a later time to share their Instagram and Facebook data through the participants unique data sharing url. If there are technical difficulties with our application pulling Facebook data, we will ask participants to manually download a file of their Facebook data from their account. No passwords would be collected by the study team. If the patient has Twitter or Reddit accounts and wish to share that data as well, they will be asked to provide their Twitter handle (username) and/or Reddit handle (username).The participants IP addresses from their devices will be collected. If subjects opt and consent to share Google take-out information with us, they will download a zip file of their google searches and YouTube activity. Google take-out data is only retrospective. They will upload the zip file to the browser https:// donate.centerfordigitalhealth.upenn.edu. If subjects opt and consent to share step tracking data with our research team and the patient is eligible and has Apple Health, S Health, Garmin, or Fitbit, they will export the data from the specific step tracking platform and the patient will upload this data to https:// donate.centerfordigitalhealth.upenn.edu. Data collection manual outlines all data extraction procedures [See Appendix]. Data extracted from Facebook will include: age, gender,

language, photo, work history, education history, hometown, interests, relationship status, current city, religious or political views, number of friends, items "liked", and number and status updates. Data from Twitter will include: handle, number of followers/following, tweets, geography, time/date, profile. Data extracted from Reddit will include: Reddit posts, comments to posts, your votes "up" or "down" to posts, and the sub-reddits(forums) that youve joined. Data from Instagram will include image posts, time of posting, and image captions. Data extracted from Google takeout includes the search terms, website visited with the URL, YouTube search history, date, and Google fit that includes activities and daily aggregations. If sharing Apple Health, data extracted will include: any data stored or gathered by the Health app and any associated devices, including any Medical ID data, the native iPhone step counter and distance tracker, any data from an Apple Watch, and any data gathered from anythird partydevices that are syncing to Health app, like a smart scale or blood pressure monitor. (http://osxdaily.com/2019/05/20/export-health-data-from-iphone/). If sharing Fitbit, data extracted will include: Activity, Exercise, Social, Sleep, Coach, Corporate, Logs, Profile, Direct messages, Female health, Sleep score, Friends, Subscriptions. If sharing Garmin, data extracted will include: steps per month for the past 12 months. If sharing S Health, data extracted will include: Daily step count for 34 days prior to export date(https://medium.com/@dimshik100/how-to-extract-your-personal-samsung-health-data-514bbe2331f7). If sharing MapMyFitness, data extracted will include: Date Submitted, Workout Date, Activity Type, Calories Burned, Distance, Workout Time, Average Pace, Max Pace, Average Speed, Max Speed, Average Heart, Rate, Steps, Source, Link *we do not use this as it requires user login to view link results* Patients who donated at least one of the following digital data sources: social media, google takeout zip file, or step tracking data will receive $50 for survey completion. For in-person enrollment, the entire process will take place during one encounter lasting approximately 15-60 minutes. For remote enrollment, the process can take 15-60 minutes in one session or up to a week or more depending on completion of all enrollment steps. If participants are unable to donate desired digital data sources, we will contact them up to 4 times in a month to download and donate all relevant sources as well as complete the baseline survey if they havent already done so. All patient enrollment status tracking will be documented and saved in either our Penn Medicine secure shared drive, Penn Box (HIPAA compliant feature), or REDcap. From the above cohort we will additionally consent these patients to complete several surveys, if they opt to participate in this additional part of the study. These include: perceptions of health status and perceptions about social media will be assessed as this could impact engagement with a digital intervention. In addition to recontacting the above cohort, we will also identify an additional cohort from iConnect (Penn Medicine research registry), who opt to participate in focus groups that will evaluate perceptions about use and engagement phenotypes, details about individual use of social media platforms, concerns and ways to mitigate concerns about phenotyping, potential uses of phenotypes, opportunities for patient centered design and implementation. Patients will complete a separate informed consent to participate and participation will be anonymous. Interviews will occur via virtual conference call and recorded. After the patient has completed enrollment, we will send them a follow-up summary about the status of the study (ie. how many participants are enrolled and the number of different data sources shared. This application and methodology have been previously used by investigators in this protocol for research within the World Well Being Project at the Positive Psychology Center at the University of Pennsylvania (prior Penn IRB protocols #816091, 813820, 819442, 824493).

Study duration

It will take approximately 15-60 minutes during one session, but could take up to a week or more depending on completion of all enrollment steps for a participant to complete the study. For remote enrollment, it could take up to a week or more for participants to complete the eligibility survey, baseline survey, and share their digital data. If there are technical difficulties with our data collection application, we will ask participants who are opting to share Instagram if we can record their Instagram username for manual data extraction. Additionally, we will reach back out to them at a later time to share through our data collection application if there were any technical difficulties. All recruitment and follow up emails will be A/B tested, also known as split testing, the process of comparing two identical versions of an email. Potential variations include the order of information, email headers, and signatures. We aim to recruit 1000 participants over 4 years. Additionally, participants who agree to participate in the optional survey and interview will have to wait until a date and time is determined for the focus group to occur. Follow-up is necessary in this case.

Resources necessary for human research protection

Describe research staff and justify that the staff are adequate in number and qualifications to conduct the research. Describe how you will ensure that all staff assisting with the research are adequately

informed about the protocol and their research related duties. Please allow adequate time for the researchers to conduct and complete the research. Please confirm that there are adequate facilities for the research.
The study team will not have direct access to any participants password for any digital media accounts and will not have the ability to post any messages or content on behalf of the user. If there are issues with the data collection application, we will ask participants to share their Instagram username with us for manual data extraction. Participants will receive a copy of consent-related documents as well as detailed information about what data are being extracted, how those data will be used and shared among the research team and instructions on how to permanently delete any study apps that provide ongoing access. We will observe best practice protocols for information security within our study team, practices that are already well-established for minimum-necessary access from environmentally, physically, and digitally secured repositories. Digital data will be reported only in aggregate and not the level of an individual, aside from excerpts of illustrative postings in the same way written questionnaires are reported only in aggregate with the exception of excerpts of illustrative free text comments. The study team will not communicate with participants via social media or publicly post that the participant is involved in the study. Participants will be informed that they should not have any expectation of surveillance or medical advice/medical response or that any posts, images, searches, physical activity data will be responded to in real time or at any time by the study team. Lexical analysis will be completed by the coinvestigators listed, whom have prior experience analyzing social media data and using the Facebook application used in this study to extract language, photos, and other social media data from Facebook, as well as applying the described lexical analysis. All aspects of the study will be overseen by the principal investigator, Raina Merchant, MD MSHP, assistant professor in Emergency Medicine, who has extensive experience analyzing social media data and large data analysis. Together this research team is part of the Penn Medicine Center for Digital Health at the Penn Medicine Center for Health Care Innovation. Directed by Dr. Merchant, this multidisciplinary team has a rich history of collaboration and we believe this is the perfect environment to successfully apply the described protocol.

Characteristics of the Study Population

Target population
We will identify patients ages 30-74 with and without CVD) who receive care in the University of Pennsylvania Health System or in another health system in the United States (U.S.). For patients with CVD, we will be using the following ICD codes for identification :ICD 10: I63, I20-I25. For patients without CVD, please find attached the list of ICD 10 codes we will use for identification

Subjects enrolled by Penn Researchers
1000

Subjects enrolled by Collaborating Researchers
0

Accrual
Patients will be contacted via email/phone call to participate in the study or in person during an inpatient admission from the Emergency Department at HUP, PAH, or PPMC or an out-patient visit at Penn Medicine Primary care clinics, Internal Medicine, Cardiology clinics or if they are a patient in another health system, we will reach out to them via email or phone call

Key inclusion criteria
- 30-74 years of age - Willing to sign informed consent - Primarily English speaking (for language analysis) - Has an account on any of the following digital data platforms (Facebook, Instagram, Twitter Reddit, Google (gmail), or smartphone or wearable device such as Apple Health, Fitbit, Samsung Health, MapMyFitness or Garmin) and willing to share data - If has social media account, Instagram or Facebook, willing to share historical and prospective data (60 days) If has Google (gmail) account, willing to download and share google takeout zip file - If has smartphone or wearable device, willing to share step data - Willing to share access to medical health records - Willing to share healthcare insurance information For the Objective 1 subgroup: - 30-74 years of age - Willing to sign informed

consent Willing to complete an online survey

Key exclusion criteria
- Patient does not meet age inclusion criteria above - Does not use and post on digital data sources we are studying or unwilling to donate data - Patient is in severe distress, e.g. respiratory, physical, or emotional distress - Patient is intoxicated, unconscious, or unable to appropriately respond to questions - Patient has a history of heart attack, stroke, or CAD more than 5 years ago. For the Objective 1 subgroup: - Not between 30-74 years of age - Unwilling to sign informed consent Unwilling to complete an online survey

Vulnerable Populations

	Children Form
	Pregnant women (if the study procedures may affect the condition of the pregnant woman or fetus) Form
	Fetuses and/or Neonates Form
	Prisoners Form
	Other
x	**None of the above populations are included in the research study**

The following documents are currently attached to this item:

There are no documents attached for this item.

Populations vulnerable to undue influence or coercion
This study will not intentionally enroll any vulnerable populations, however pregnant women may be enrolled if they meet eligibility requirements as described above. All the research assistants have completed the CITI-Protection of Human Subjects Research Training courses which fulfill the University's requirement for training in human research protections. RAs will ensure that subjects' understand that participation in this study is voluntary and will not effect their care in any way

Participant recruitment
Please describe the plan to fairly identify and recruit a group of participants that is reflective of the population under study. If this is a multicenter protocol, the recruitment plan should describe the local (Penn) site's plan. Describe: how potential participants may be identified (review of medical records, Slicer Dicer, DAC reports including referrals from physician offices and clinics);who may approach potential participants;methods to achieve a sample representative of the population being studied;what information may be presented to or discussed with them; andthe context and setting in which recruitment will happen.
Recruitment will occur in-person and remotely. For in-person recruitment, patients will be directly approached by a research assistant (RA) or AAs. For remote recruitment, patients will be reached out to via email or over the phone by a study team member. For in-person recruitment, if necessary AAs or RAs will communicate directly with patients providers (physicians, residents or nurses) to determine whether it is appropriate to approach the patient. For remote recruitment, RAs and AAs will conduct a brief chart review to ensure that the patient is able to consent to participate in studies. Potentially eligible patients will be approached or reached out to remotely to verify they meet all inclusion criteria and will then be asked to participate. An AA or RA will approach or call/email/MPM/W2H participants and explain the study in brief. Interested participants may complete the informed consent online via REDcap or https://donate.centerfordigitalhealth.upenn.edu/capture/ process at that time, or ask the research assistant to return or call at a, more convenient time. After informed consent, the participant will be asked to share the digital data types that they use (Facebook, Instagram, Twitter, Google search, step data). Finally, participants will complete the survey. All recruitment and follow up emails will be A/B tested, also known as split testing, the process of comparing two identical versions of an email. Potential variations include the order of information, email headers, and signatures. For patients that opt to participate in the focus group, they will receive a date and time to for the virtual conference call and it is a one-time occurrence. Additionally, we will use flyers and post our study on the University of Pennsylvanias research registry (iConnect) for interested patients to reach out to us regarding participation in our study. Within iConnect, we will use their pre-screener feature to assess patients for pre-eligibility. These questions will be the same questions as we have already built in our eligibility

survey. For interested patients reaching out about the research study, our RAs and AAs will assess for eligibility and reach back out to the patient if eligible or ineligible. If patients are eligible and consent to participate, RAs and AAs will start the enrollment process. We will also post our study on Penn Medicine Lancaster General Healths various social media sites (ie. Facebook, Instagram, Twitter, LinkedIn, etc.) with a link to our iConnect study site. For interested patients reaching out about the research study, our RAs and AAs will assess for eligibility and reach back out to the patient if eligible or ineligible. If patients are eligible and consent to participate, RAs and AAs will start the enrollment process. Lastly, we will be using Research Matchs (https://www.researchmatch.org/) site to reach out to interested individuals. For interested patients reaching out about the research study, our RAs and AAs will assess for eligibility and reach back out to the patient if eligible or ineligible. If patients are eligible and consent to participate, RAs and AAs will start the enrollment process. After the patient has completed enrollment, we will send them a follow-up summary about the status of the study (ie. how many participants are enrolled and the number of different data sources shared. We recontact study participants and offer them the opportunity to complete 1 additional survey and opt-in interview. In addition to recontacting study participants, we will recruit participants (specific to Objective 1) using ResearchMatch and iConnect registries.

Recruitment Materials
Is the research team using any recruitment materials? These may include but are not limited to: phone call scripts, radio/video scripts, flyers/brochures, internet postings, email, letters to potential participants, letters to patient physicians, My Penn Medicine (MPM), other direct messaging, etc. For guidance regarding recruitment materials, please review the IRB's guidance on Participant Recruitment Materials online:https://irb.upenn.edu/recruitment
Yes

Use of Penn Media & Social Media Services
Will the recruitment plan propose to use any Penn media services (communications, marketing, etc.) for outreach via social media avenues (examples include: Facebook, Twitter, blogging, texting, etc.) or does the study team plan to directly use social media to recruit for the research?
Yes

Please identify which method(s) of social media you will utilize, the content of the text to be used, and the method(s) for posting this information (i.e., using Penn supported communication services). When proposing the text to utilize, please be aware of any social media limitations (i.e., number of characters allowed in a tweet) and any appropriate confidentiality practices necessary to be compliant with posting research recruitment text.NOTE: Penn Medicine must utilize one of the centralized PM Facebook Pages:ClinicalResearch@Penn Facebook pagePenn Medicine Facebook page@PennCancer Facebook pageAll clinical research paid Facebook ads must be listed on Clinical Research @ Penn Facebook page,www.facebook.com/ClinicalTrialsAtPenn. Exceptions to the above must get approval from the Penn Medicine Social Media Committee: pennmedicinesocialmediacommittee@uphs.upenn.edu.
This study uses the Penn Medicine iConnect Facebook page to recruit participants.

The following documents are currently attached to this item:

There are no documents attached for this item.

Subject compensation*
Will subjects be financially compensated for their participation?
Yes

The following documents are currently attached to this item:

There are no documents attached for this item.

If there is subject compensation, provide the schedule for compensation per study visit or session and total amount for entire participation, either as text or separate document

Patients who consent to share digital data from 1. Social media, 2. Online search, and/or 3. Step data) 4. Electronic medical record, and 5. Their health insurance information will receive $50 for completing the survey and data donation. Patients who consent to complete additional surveys will receive an extra

$10. Participants who complete the additional survey will receive $20 compensation and $30 for interview completion. All payments will be loaded on a Virtual Visa Greenphire ClinCard. Patients who consent will be entered into a raffle to win an iPad or equivalent device valued at approximately $500. Raffle winners will be chosen at random for every 300 study participants enrolled and notified by email and/or phone

Study Procedures

Suicidal Ideation and Behavior
Does this research qualify as a clinical investigation that will utilize a test article (ie- drug or biological) which may carry a potential for central nervous system (CNS) effect(s)? Centeral nervous system(CNS) effect: the ability of a test article to enter into and potentially interact with the central nervoous system (brain and spinal cord). Clinical Investigation: Any experiment that involves a test article and one or more human subjects that either is subject to requirements for prior submission to the Food and Drug Adminstration (FDA) under section 505(i) or 520(g) of the Federal Food, Drug, and Cosmetic Act, or is not subject to the requirements for prior submission to the FDA under these sections of the act, but, the results of which are intended to be submitted later to, or held for inspection by, the FDA as part of an application for a research or marketing permit.
No

Procedures
The study will be completed within 1-5 encounters when the participant is at UPHS or remote. Should the encounter be interrupted before participants share their social media or digital data information, or should participants forget social media login credentials during the encounter, they will receive up to 5 emails within 1-2 weeks of consenting reminding them to log in and share their social media data remotely on https:// donate.centerfordigitalhealth.upenn.edu. Each participant will have a unique ID. For example 1304, is the unique ID associated with a participant, (https:// donate.centerfordigitalhealth.upenn.edu/socialshare?activity=1304&survey=69) For patients that opt to participate in the focus group, they will receive a date and time to for the virtual conference call and it is a one-time occurrence. After the patient has completed enrollment, we will send them a follow-up summary about the status of the study (ie. how many participants are enrolled and the number of different data sources shared.

The following documents are currently attached to this item:

There are no documents attached for this item.

Deception
Does your project use deception? Deception could be considered any direct misinformation presented to the subject or omission of key information pertaining to the design or nature of the project.
No

International Research
Are you conducting research outside of the United States?
No

Analysis Plan
As digital data exist in multiple formats (e.g. posts, photos, searches, steps) we will analyze each source according to its specific attributes. We will systematically apply robust data-driven approaches to parsing and stratifying data. This step requires significant data pre-processing and annotation as raw digital data is not directly interpretable. This will also occur over multiple iterations over the course of the grant as new terms (e.g., the term "vaping" describes use of e-cigarettes) and digital sources emerge. To automatically extract meaningful variables from the text, we will use both dictionary-based (e.g., counting words from previously constructed word lists, lexica) and data-driven "open vocabulary" methods which automatically generate lexica and identify important words and phrases based on the data itself rather than a priori categories. For dictionary analyses, we will use the Consumer Health Vocabulary, an online open source thesaurus of colloquial and technical medical terms and the Unified Medical Language System which contains lists of medical terms and vocabulary to identify disease

discussions. We will specifically focus on words and words associated with previously identified predictors of CHD (e.g. demographics, family history, lifestyle, comorbidities, blood pressure) as identified in a systematic review of prediction models for cardiovascular risk by Damen et al. BMJ 2016. We will create classifiers to determine if a given instance of a disease term is referring to the disease or something unrelated (e.g. "If I see a celebrity I will have a heart attack" vs. "I am having chest pain, SOB and had a prior heart attack"). Our prior work has demonstrated that, as one might guess, some CV diseases, like "hypertension", are almost always used in their medical context while others, like "heart attack", are often used colloquially. We will also create classifiers to categorize language by broad grouping such as risk factors, treatment, medication adherence, symptoms, and by more granular categories such as active/sedentary, healthy food/unhealthy food, smoking. We will compare the performance of several machine-learning algorithms (e.g. support vector machines, random forests, and, in cases with adequate data, deep learning) and assess the accuracy of our disease term classifier against hand-annotated data. For multiword phrases we will use approaches for determining the probability that words occur together (e.g., pointwise mutual information PMI). We will also use additional linguistic models to account for more subtle language complexities and nuance such as parts of speech, and disambiguated words. We will use automated computational methods (e.g., deep learning-based object recognition) combined with manual coding to classify features (e.g., healthy or unhealthy food, physical activity, smoking, and drinking) in images. To determine nutritional content of food images, for example, we will use the imagga.com image classification software, combined with analysis of the words used in the text associated with images to recognize food images. Those images that are food images will then, to the best of our ability, be compared to a list of healthy and unhealthy groupings (e.g. high salt, high fat) and also compared to a USDA database of precise nutritional values for over 30 nutrients for 8600 food items. We will also construct a convolutional neural network based on labeled food and health behavior-related images to train a model to recognize images in the most frequent categories which our first pass fails to find. We will analyze the labeled images to better understand the nature of food posting, and how they vary with the time of day, demographics of the person posting the image We will compare summary data from smartphones with self-reported data. Time series data from sensors will be used to analyze presumed changes in behavior using a previously validated health behavior activity change detection framework which segments time series data by time periods, identifies changes and their significance, and identifies the presumed locus of changes. The strength of using different types of data (text, images, sensors) is that each provides a unique type of information. Sensors provide direct measurements of protective and risk factors for CV, but they only measure what they are intended to measure. On the other hand, text and image data have been shown to capture a wide variety of information about people from demographics and personality to mood and beliefs. Further, each patient may use differing proportions of each modality such that, for example, text may make up for a lack of images and vice-versa. As digital media data may over-represent, under-represent or not represent health behaviors we will compare survey data and clinical data with digital data to quantify differences in behaviors and risk factors. We will explore statistical and machine learning methods for combining heterogeneous variables into predictive models. For our main analysis, we will both explore the correlations between the different modalities to better understand when they provide redundant or non-redundant information and we will, more importantly, combine the different modalities into a single predictive model. Standard machine learning methods such as random forests make it easy to add heterogeneous features to a single predictive model, and give measures of variable significance' indicating how much each feature contributes to predictive accuracy. We will also test more sophisticated machine learning methods for combining multiple modalities in a single model, such as using extensions of the 'elastic net' regression and dimensionality reduction framework that account for the fact that different modalities require different regularization penalties. We will also explore imputation and "back-off" methods that account for the fact that different patients will have substantially different amounts of data available in different modalities. These methods work by either estimating ("imputing") values for the missing data or automatically adjusting the weights given to different modalities for each patient based on the amount of data available for that patient. For Objective 1: Sample characterization: We will use summary statistics to describe patient demographics, survey response data, and digital data usage (e.g. frequency of posting, number of social media platforms). Descriptive data will be presented as mean (SD) for continuous variables and frequencies of participants for categorical variables. 5b Extracting topics and features: We will use the Mallet package for implementation of the Latent Dirichlet Allocation (LDA) method for clustering language data of all participants. The LDA probabilistic model assumes that documents (e.g. Facebook status updates) contain a distribution of topics which then contain a distribution of words. Ultimately, words are grouped together by considering the other words they appear with. Descriptive statistics, t-tests (continuous measures) and chi-squared tests (categorical measures) will be used to compare topics and

features with survey data. For example, NHANES survey questions (e.g. Do you now smoke cigarettes?) would be compared with images of cigarette smoking. For Objective 2: We will compare prediction models with and without digital variables. Probability weighted Cox proportional hazard analysis with robust variance estimates will be used to assess the association between each variable from digital media and CHD in univariable and multivariable models. We will use graphical and analytical methods to test the proportional hazards assumption for each analysis. To evaluate the predictive strength of digital data, we will build four predictive machine learning models using a stepwise sequence to select the most parsimonious model using combinations of risk factors. (1. Framingham predictors / American College of Cardiology ASCVD Risk Estimator Plus alone, 2. Framingham predictors / American College of Cardiology ASCVD Risk Estimator Plus + SES, 3. Framingham predictors / American College of Cardiology ASCVD Risk Estimator Plus + digital topics/ features, 4. Framingham predictors / American College of Cardiology ASCVD Risk Estimator Plus + digital topics/features+ SES) For Objective 3: We will use Super Learner, the ensemble machine learning approach which allows for the specification of multiple plausible candidate prediction models. Each of the candidate models is applied to a training set and outcomes are predicted using the validation set. A loss function is calculated within each validation set and then averaged across validation sets, which provides the estimated cross-validated risk score for each method. Calibration will be assessed using a calibration plot comparing the predicted probabilities (obtained using fixed regression coefficients of the prediction models) with the observed probabilities. The Super Learner algorithm finds the optimal weighted combination across all of the specified methods. Van der Laan et al. proved the asymptotic efficiency of the Super Learner algorithm and demonstrated that the optimal combination performs at least as well as the best estimators from the candidate models. We will implement Super Learner in R (The R Foundation for Statistical Computing). In this approach generalized estimating equation (GEE) and generalized linear mixed effects models will be explored to account for repeated measures per patient over time. One of the challenges of cost analyses is censoring, since cost may be incomplete and available for some patients and not others. As the dependent variable will be cost, we will be attentive to informative censoring and right skewness, potentially using generalized linear models under a gamma distribution and log link. To account for informative sensoring we will use inverse probability of sensoring weighting. We will also explore the application of our new nested g-computation approach that allows for informative censoring and time-varying covariates to obtain future cost predictions. Additionally, we will generate summary statistics to describe the cohort and survey responses. We will use the model-based approach of latent class analysis (LCA)- to classify individuals into previously unmeasured subgroups. Variables to include in the LCA will include: demographics, years of platform use, posting: volume, quality, retransmission, and engagement. We will use the Bayesian information criterion to assess goodness of model fit, entropy to assess variations between classes, and the parametric bootstrapped likelihood ratio test to assess if a model with k classes has better performance than k-1 classes. We will then use a regression model to compare with measures of health status, social media perceptions, and readiness/activation for each latent class. Models will be adjusted for demographics.

The following documents are currently attached to this item:

There are no documents attached for this item.

Data confidentiality

- x **Paper-based records will be kept in a secure location and only be accessible to personnel involved in the study.**
- x **Computer-based files will only be made available to personnel involved in the study through the use of access privileges and passwords.**

  Prior to access to any study-related information, personnel will be required to sign statements agreeing to protect the security and confidentiality of identifiable information.
- x **Wherever feasible, identifiers will be removed from study-related information.**

  A Certificate of Confidentiality will be obtained, because the research could place the subject at risk of criminal or civil liability or cause damage to the subject's financial standing, employability, or liability.

  A waiver of documentation of consent is being requested, because the only link between the subject and the study would be the consent document and the primary risk is a breach of confidentiality. (This is not an option for FDA-regulated research.)
- x **Precautions are in place to ensure the data is secure by using passwords and encryption, because the research involves web-based surveys.**
- x **Audio and/or video recordings will be transcribed and then destroyed to eliminate audible identification of subjects.**

Subject Confidentiality

Study personnel will have access to patients' electronic medical records for screening, eligibility confirmation purposes. Study personnel will also have access to patients past medical history and demographic data. The information used for screening or enrollment purposes and study analysis. This data will be extracted using Clarity. On the servers that host the Facebook language-extraction application, we will not collect identifiable information. Instead, we will use the Facebook/Instagram mechanism to generate a unique random ID. This ID will be used to track users and will be stored together with their survey results on the client-facing server. If disclosed, the unique random ID cannot be traced back to a given individual (unlike the Facebook/Instagram user ids, which can comparatively simply be resolved into identifiable information). In addition to this user-facing server (application server), we will also have a Secure Data Repository, which will receive the Facebook user id from the application server and will then proceed to query the Facebook/Instagram Application Programming Interface (API) for (identifiable) user information and Facebook/Instagram statuses the users have written on Facebook/Instagram. This we can only do once the user has joined our application, these privileges are managed through Facebook's/Instagram's systems of access tokens. The Secure and the Application server would use the same access token, but only the Secure Server would pull identifiable user information, whereas only the Application server would interact directly with the users (and is thus more at risk for code injection and other hacking attacks). The secure data repository will be housed on servers under our control at the University of Pennsylvania, such that no data we collect will be stored out of our control. All language data will be stored on the Health Service Research Data Center (HSRDC). As per the HSRDC Data Management Policy and Procedures Manual, password protection is used at the server and web portal levels for all transactions that allow entry and editing of data, or provide access to sensitive subject data or administrative privileges. Passwords will be managed to require all users to change their password within 90 days and strict rules will be implemented to require strong passwords. Additionally, all PHI data will be encrypted and linked to survey data through a study ID number, with linkage and de-encryption keys activated only by a user password for which a member of the research team has been given permission to access these sensitive data (Principal Investigator Dr. Merchant and project staff who are under the direct supervision of the PI). As per the HSRDC Data Management Policy and Procedures Manual, the database management on these servers is built with multiple layers of security and follows best practices for securing sensitive data. The main levels of security are fourfold and include machine physical security at the IT facility at which the servers are housed, physical server security in the highly restricted server containment room, electronic server security (firewalls, passwords, encryption) restricting access to the machine, and directory access controls restricting access to these particular data. We will consult the PI, ED physician to manage participant safety concerns and serve as the clinical contact point person. The data security of the applications is as follows: Facebooks General Data Protection Regulation - https://www.facebook.com/business/gdpr Instagram Data privacy webpage, https://help.instagram.com/519522125107875 Google

takeouts security center - https://www.google.com/safetycenter/everyone/start/security-help/, Reddits Data protection webpage, https://www.reddit.com/r/dataprotection/.We would report to the specific digital platform anything that violates their community standards. We will also apply existing filters (e.g. not safe for work) before processing image and text data. Facebook community standards: https://www.facebook.com/communitystandards Instagram community standards: https://help.instagram.com/477434105621119 Google takeout standards, https://www.google.com/+/policy/content.html. Our data scientist will manage the collection of these social media platforms. Only our research team will have access to the data through the applications. For the experimental group, the patient and provider will seem summary trends from the social media platforms. If study participants do not uninstall the app after the study ended, the apps will be closed and no additional information will be used or collected through the app. All patient information used for screening, recruitment, and enrollment purposes, survey responses, and focus group input will be stored on either the Penn Medicine secure shared drive, Penn Box (HIPAA compliant feature, or REDcap.

Sensitive Research Information*

Does this research involve collection of sensitive information about the subjects that should be excluded from the electronic medical record? [NOTE: This does not apply to: 1) research information that would not normally be included in the electronic medical record or 2) information that is in the electronic medical record as part of clinical care.]

No

Subject Privacy

Privacy refers to the person's desire to control access of others to themselves. Privacy concerns people, whereas confidentiality concerns data. Describe the strategies to protect privacy giving consideration to the following: The degree to which privacy can be expected in the proposed research and the safeguards that will be put into place to respect those boundaries. The methods used to identify and contact potential participants. The settings in which an individual will be interacting with an investigator. The privacy guidelines developed by relevant professions, professional associations and scholarly disciplines (e.g., psychiatry, genetic counseling, oral history, anthropology, psychology).

All encounters will take place in private rooms, via the online platform, via email communication, via secure text messaging, or through a secure phone call. We will offer patients the option of having the survey completed without any other people present in their room or they can complete it on their own time. Patients have the option of the research assistants assisting them in completing the survey or filling out the survey by themselves. Patients who choose to share their Facebook, Twitter, Instagram or other digital data will be given a unique ID that will be linked to identifiable information. Participants who install the application to share their Facebook posts will be able to opt in and out of the application at their discretion. Patients will be asked by a RA if they would be willing to anonymously share their Facebook/Instagram posts for health research (Example description: "Penn researchers are trying to learn if certain language is associate with certain health outcomes. Would you be willing to share your Facebook posts with researchers? Your posts will be anonymous -- that is, not connected to you name -- and you can choose to stop sharing your posts at any time.") The Facebook/Instagram application will access data that is stored in the user profiles in the Facebook/Instagram platform. When participants first click on the link to use the app, they will first be greeted by the Facebook/Instagram consent screen that will ask the user permission for the application to have access to their profile information and to consent to the terms and service of the application. Participants would have to agree to this screen in order to use the application. They will have to click agree to access the application (at least at the first time). All language data will be stored on secure severe using only the unique study ID which can be used to link this data to Center for Digital Health data collection platform, (https://donate.centerfordigitalhealth.upenn.edu/capture/). This linkage will be stored in a double locked fashion on a password protected computer will access only by the principal investigator and key study personnel. For the participants who select to enroll in an interview, interview audio-recordings will be transcribed verbatim, and all identifiers will be removed. Transcripts and audio files will be saved on a Penn Medicine managed, HIPAA compliant server. Interview audio-recordings will be transcribed verbatim and all identifiers will be removed. The recording and transcript will be kept a secure and locked area with access limited to designated researchers. After we analyze the recordings, we will destroy recordings after data analysis or completion of the study. All encounters will take place on Penn Medicine Zoom videoconference. When utilizing Zoom for audio recordings, additional steps will be made by the research team to ensure participant safety. Based on guidelines, researchers will ensure the following settings for privacy and compliance for Penn Zoom. 1) Researchers will turn off the Personal Meeting ID setting and instead opt for the meeting ID to be generated automatically, allowing for new

links and meeting IDs to be created for each meeting. 2) Research staff will turn off Cloud Recordings and ensure only local recordings that are HIPAA compliant are used as they are encrypted and managed by PMACS/DART or UPHS. The recording and transcript will be kept a secure and locked area with access limited to designated researchers. After we analyze the recordings, we will destroy recordings after data analysis or completion of the study. All survey data will be stored in a secure, web-based database (Way to Health or REDCap) where a study ID number will be generated for each patient. A link between the study ID number and the patient PHI will need to be maintained to ensure that the study staff can track recruitment efforts to potential participants and to avoid contacting any patients who have previously declined to participate. To ensure that patient confidentiality is preserved, individual identifiers (such as name) are stored in a single password protected system that is accessible to study research, analysis and IT staff only. This system is hosted on site at UPenn and is protected by a secure firewall. Once a participant is in this system, they will be given a unique study ID number. Any datasets and computer files that leave the firewall will be stripped of all identifiers besides the study ID and individuals will be referred to by their study ID only. The study ID will also be used on all analytical files. Way to Health and REDCap are secure web applications for building and managing online surveys and databases. Privacy of all study data will be maintained by restricting access to the identifiable information only to approved study staff who have received subject confidentiality and privacy training. Study coordinators will access patient contact information from the database to conduct recruitment phone calls. The study coordinator will review the consent script, which will include a description of the voluntary nature of participation, the study procedures, risks and potential benefits in detail. Participants will be told that all information will be kept strictly confidential, except as required by law. Subjects will be provided a copy of the consent document. All efforts will be made by study staff to ensure subject privacy. This database is hosted on a secure server as detailed in the subject confidentiality section.

Disclosures
Will any data or specimens from Penn participants OR other research generated work product (e.g., intellectual property) be disclosed to any individuals, entities, or vendors, etc. outside of Penn?
No

*Data Protection**

- x **Name**
- x **Street address, city, county, precinct, zip code, and equivalent geocodes**
- **All elements of dates (except year) for dates directly related to an individual and all ages over 89**
- x **Telephone and fax number**
- x **Electronic mail addresses**
- x **Social security numbers**
- x **Medical record numbers**
- x **Health plan ID numbers**
- **Account numbers**
- **Certificate/license numbers**
- **Vehicle identifiers and serial numbers, including license plate numbers**
- **Device identifiers/serial numbers**
- x **Web addresses (URLs)**
- x **Internet IP addresses**
- **Biometric identifiers, incl. finger and voice prints**
- x **Full face photographic images and any comparable images**
- x **Any other unique identifying number, characteristic, or code**
- **None**

Does your research request both a waiver of HIPAA authorization for collection of patient information and involve providing Protected Health Information ("PHI") that is classified as a "limited data set"

(city/town/state/zip code, dates except year, ages less than 90 or aggregate report for over 90) to a recipient outside of the University of Pennsylvania covered entity?
No

Tissue Specimens Obtained as Part of Research*
Are Tissue Specimens being obtained for research?
No

Tissue Specimens - Collected during regular care*
Will tissue specimens be collected during regulator clinical care (for treatment or diagnosis)?
No

Tissue Specimens - otherwise discarded*
Would specimens otherwise be discarded?
No

Tissue Specimens - publicly available*
Will tissue specimens be publicly available?
No

Tissue Specimens - Collected as part of research protocol*
Will tissue specimens be collected as part of the research protocol?
No

Tissue Specimens - Banking of blood, tissue etc. for future use*
Does research involve banking of blood, tissue, etc. for future use?
No

Genetic testing
If genetic testing is involved, describe the nature of the tests, including if the testing is predicative or exploratory in nature. If predictive, please describe plan for disclosing results to subjects and provision of genetic counseling. Describe how subject confidentiality will be protected Note: If no genetic testing is to be obtained, write: "Not applicable."
Not applicable

Consent

1. Consent Process

Overview
We will enroll patients in person and remotely. For in-person enrollment, we will meet patients during their in-patient stays at HUP, PAH, or PPMC or an outpatient primary care visit at Penn Medicine who are between the ages of 30-74. For remote enrollment, we will call/email/text message potentially eligible patients. Potentially eligible patients will be approached or called/emailed/text messaged to verify they meet all inclusion criteria and will then be asked to participate. An AA or CDH RA will first ask participants if they use Social Media, have a Gmail account, and what type of step data. For those who do not use any, the encounter ends as these patients are not eligible. Patients who are potentially eligible, they will be told that they may be eligible for a research study looking at digital data and coronary heart disease (CHD). Subjects will be asked to confirm whether they will share their digital data with health researchers as described in this protocol. The language will be similar to (but not verbatim): "Penn researchers are trying to learn if digital data is associated with cardiovascular health. Would you be willing to share your Facebook posts with researchers? Your posts will be anonymous -- that is, not connected to your name -- and you can choose to stop sharing your posts at any time." The RA will obtain informed consent using the attached consent form via a HIPAA-compliant online survey/data collection platform (https://donate.centerfordigitalhealth.upenn.edu/capture/), or REDcap. The participants will sign an electronic version of the consent form (stored on the Center for Digital Health data collection platform,(https://donate.centerfordigitalhealth.upenn.edu/capture/) or REDcap and will be either emailed a copy of their signed informed consent or handed a copy of the informed

consent for their records. The RA or AA will describe the study in detail, including the use of information from their medical record, the survey, the use of Google takeout, how their steps are calculated, and the use of the Facebook/Instagram app to extract their posts for language analysis and how Reddit and Twitter data is extracted. The participant may refuse to answer any questions for any reason, can skip any question, and can stop participation at any time. Subjects may contact the research team as described in the consent to have their information destroyed. The research presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context. All study activities are conducted either while the patient is the in the hospital during this brief encounter, over the phone using a HIPAA secure telemedicine application (Doximity), via email using our Centers shared pennmedicine.upenn.edu email address, or via MPM or W2H. The research presents no more than minimal risk of harm to subjects and does not involve procedures for which written consent is normally required outside of the research context. The study is not intervening or currently changing patient care. Participants who complete the additional survey will complete a separate informed consent document on RedCap. All participants will have the option to complete the document with a research coordinator.

Children and Adolescents
N/A

Adult Subjects Not Competent to Give Consent
N/A

2. Waiver of Consent

Waiver or Alteration of Informed Consent*
No Waiver Requested

Minimal Risk*

Impact on Subject Rights and Welfare*

Waiver Essential to Research*

Additional Information to Subjects

Written Statement of Research*
No

If no written statement will be provided, please provide justificiation

The following documents are currently attached to this item:

There are no documents attached for this item.

Risk / Benefit

Potential Study Risks
The potential risks associated with participating in the study are minimal. Regarding unforeseen safety issues, the researchers will take every step to ensure these do not occur. Participants will have every opportunity to stop the survey at any time or discontinue participation.

Potential Study Benefits
There are no direct benefits to participation. Participants are contributing to health research in general by providing information which could improve ability to use language and social media data to predict certain health outcomes and health care utilization.

Alternatives to Participation (optional)
N/A

Data and Safety Monitoring
The principal investigator will be responsible for monitoring the data of the study. The safety, privacy, and data integrity will be maintained as described in the confidentiality section. The PI will ensure all researchers are adhering to the study protocol and applicable research regulations and Penn requirements. This includes weekly lab meetings and monthly discussions of data integrity and study progression. Interim analysis will be completed regularly for the duration of the study. All digital media data, focus group, and survey data will be collected and stored on servers that sit behind the health system firewall. Subject names and other relevant PHI will be removed if data are to be shared with auditors and other investigators for whom this information is not required. Consent materials and documentation will be kept as part of the secured study files. All study data will be carefully structured and maintained such that participants will be protected. Data will be kept in locked research offices and stored on the secure Penn server. Computer files containing data will be password-protected, and records will only be made available to research personnel who are IRB approved.

The following documents are currently attached to this item:

There are no documents attached for this item.

Risk / Benefit Assessment
Based on the potential risks and benefits described above, we believe this study may be considered minimal risk.

General Attachments

The following documents are currently attached to this item:

Cover Letter (letter.docx)

Additional forms (screenshot2025-08-01at1.42.438239pm.png)

e-copy of the informed consent document you signed virtually

UNIVERSITY OF PENNSYLVANIA
RESEARCH SUBJECT
INFORMED CONSENT AND HIPAA AUTHORIZATION FORM

Protocol Title:	Using Digital Data to Predict CHD
Principal Investigator/ Emergency Contact	Raina Merchant, MD 3400 Spruce St, Philadelphia, PA 19104 215-746-8681 Raina.Merchant@pennmedicine.upenn.edu
Study Contact	Megan Bradley (215) 615-2806 Dydhearthealth@Pennmedicine.upenn.edu

Research Study Summary for Potential Subjects
You are being invited to participate in a research study. Your participation is voluntary, and you should only participate if you completely understand what the study requires and what the risks of participation are. You should ask the study team any questions you have related to participating before agreeing to join the study. If you have any questions about your rights as a human research participant at any time before, during or after participation, please contact the Institutional Review Board (IRB) at (215) 898-2614 for assistance.

The research study is being conducted to learn more about how digital data (social media, online search, mobile media) are associated with coronary heart disease (CHD) and related risk factors, and health care utilization.

If you agree to join the study, you will be asked to complete the following research procedures:
- Share data from any or all of the following digital data sources: social media data (Facebook, Instagram, and/or Twitter), Google search history, or step tracking data
- Complete a survey that asks about diet behavior, physical activity and physical fitness, smoking and tobacco use, and digital data use
- Share information about your insurance provider and share information about your medical history

Your participation could last between 15-60 minutes or up to one week or more for study enrollment, and your data will be stored during the duration of the study. If you decide to participate in the optional focus group part of the study, follow-up is necessary.

There is no direct benefit to your participation. However, your participation could help us understand how digital data and health record information can be used to predict heart health and health care use. This study has minimal risks, but the most common risks of participation include the disclosure of potentially sensitive information. Another potential risk in permitting your data to be stored in our research database is the potential risk of loss of confidentiality.

Please note that there are other factors to consider before agreeing to participate such as additional procedures, use of your personal information, costs, and other possible risks not discussed here. If you are interested in participating, a member of the study team will review the full information with you. You are free to decline or stop participation at any time during or after the initial consenting process.

Why am I being asked to volunteer?
You are being invited to participate in a research study because you are between the ages of 40 and 74 and you use and post on digital data types that we are studying.

If you decide to participate, you will be asked to sign this form. Your doctor may be an investigator in this research study. You do not have to participate in any research study offered by your doctor. If you choose not to participate, there will be no loss of benefits to which you are otherwise entitled. You may also decide to discuss the study with your family, friends, or family doctor. Being in a research study is different from being a patient. As an investigator, your doctor is interested both in your clinical welfare and in the conduct of this study.

What is the purpose of the study?
The purpose of the study is to learn more about how digital data (social media, online search, steps) can give doctors a better understanding of cardiovascular risk by providing insights about an individuals' daily habits.

Why was I asked to participate in the study?
You are being asked to join this study because you are between the ages of 40 and 74, English is your primary language, use digital data types that we are studying.

How long will I be in the study?
The study will take place over a period of 5 years, but your participation will end after you complete all the enrollment steps. The study enrollment will last approximately 15-60 minutes during one session. The enrollment process could take up to a week or more depending on completion of all enrollment steps. You will complete all enrollment steps, completing surveys and downloading digital data, online through email, over the phone, or via text messaging on your own device or a study team members' device remotely or in-person on a secure study laptop with a study team member.

What will I be asked to do?
- You will first complete a survey that takes approximately 10-15 minutes to complete. The survey will ask questions about your diet behavior and nutrition, physical activity and fitness, smoking and tobacco use, digital data use, and basic demographic information.
- After completing the survey, you will be asked to share whatever digital data sources you have decided to donate.
- If you decide to share Facebook or Instagram, you will then be directed to the study app that was designed for the purposes of this study.
    - If sharing Facebook, you will be asked to sign in to your Facebook account using our secure application. When you sign into your account, you grant access to a Facebook application called Penn Image Station that was designed for the

purposes of this study. The Penn Image Station application will gather information from your Facebook public profile, status updates, personal description, and "likes". The application will not affect your usual use of Facebook nor will it post to your page. Penn Image Station will collect historical Facebook data (for example, when you created your account) and for the duration of the study, 60 days or until you no longer wish to participate. Data extracted from Facebook may include bio (about me section) text and image posts, time of posting, pages liked, and image captions. The application will not affect your usual use of Facebook nor will it post to your page.

- OF NOTE: If there are technical difficulties with our application pulling Facebook data, we will re-contact you via email, text message or telephone call (your preference) at a later point once the application is running.

- If you share your Instagram posts, you will be asked to sign into your Instagram account on our secure application. When you sign into your account, you grant access to an Instagram application called Penn Social Mediome. You will read a brief statement, either agree or disagree with the statement, and then log on to your Instagram account. Your Instagram posts and photos will be extracted by the Penn Social Mediome application. Penn Social Mediome will collect historical Instagram data and for the duration of the study, 60 days or until you no longer wish to participate. You may notify us in writing that you no longer wish to participate in the study. Data from Instagram will include image posts, time of posting, and image captions. Instagram data will only include your past data.
- OF NOTE: If there are technical issues with our secure data collection application, we may ask you to share you Instagram username with us so we can extract the data manually.

- If you share your Twitter posts, you will provide the study team with your Twitter username or "handle." Data from Twitter will include: handle, number of followers/following, tweets, geography, time/date, profile. Your Twitter data will be extracted and de-identified by researchers. Twitter data will be collected for the duration of the study (5 years) or until you no longer wish to participate
- If you share your Reddit posts, you will provide your Reddit username/handle with the study team. Your Reddit posts comments to posts, their votes "up" or "down" to posts, and the sub-reddits (forums) that they've joined will be extracted and de-identified by researchers. Reddit data will be collected for the duration of the study (5 years) or until you no longer wish to participate
- If you decide to share your Google takeout (online searches and tracking information) data with us, you will receive directions either in-person or via email from one of our team members. You will receive a document explaining how to download your data through Google take-out. Please note, the study will not affect your Google use. Your search history, YouTube histories, and other activities will be only retrospective. Data extracted from Google takeout includes the search terms, website visited with the URL, YouTube search history, date, and Google fit that includes activities and daily aggregations.

- If you decide to share your daily step count with us. This can be done different ways depending on what step-tracking platform you have an account on. The research assistant will guide you through this process.
- As part of the study, we will also collect health information from your electronic medical record to be used for research purposes. Data from the medical record will include but not be limited to: demographic information (e.g. date of birth, insurance type, height, weight), medical diagnoses (e.g. history of hypertension, diabetes, cancer, depression), mode of arrival to the hospital (e.g. ambulance or other), labs (e.g. cholesterol, lipid panel, HbA1c), vital signs, mediations, procedures, dates and diagnoses from outpatient visits, emergency room visits, and hospitalizations. Your authorization for your personal health information for this specific study will last the duration of the study. You may notify us in writing that you no longer wish to participate in the study. If you are not a patient within the University of Pennsylvania Health System, also known as Penn Medicine, we will ask you to upload screenshots and/or self-report information from their office visit summaries, labs, vital signs, ED and inpatient hospital admission information since we will not have access to your electronic medical record.
- We are asking permission to collect information on your Medicare and/or insurance coverage and on your cost of care. This information will be collected directly from your insurance and medical record. Information will be collected for up to 5 years. If you agree to participate in the study, the following information will be collected in regards to your health insurance: type of health insurance, health insurance provider, health insurance policy, health insurance group number, and the policy holder's name and date of birth will be collected. OF NOTE: If we aren't able to obtain your health insurance policy number, we will either call you to obtain your policy number or you can fill out a secure survey through REDcap that will allow you to input your policy number.
- If you decide to participate in the optional additional survey part of the study, we will ask you questions related to your health status perceptions, perceptions of social media use, and engagement on social media.
- If you decide to participate in the optional focus group part of the study, we will ask you to virtually join a group of 15-20 patients via conference video or phone call. You will receive a date and time with video conference login information to join the focus group session. The focus group is a one-time occurrence. The focus group will be recorded. All recordings will be transcribed and the original recording will be erased to protect your identity.

For this study we may need to contact you via email to provide you information about scheduling or send you information about your participation in the study. Email communications are often not secure and may be seen by others as a result. By signing below, you accept this risk. If you wish for us to use a different means to communicate with you during the course of this trial please discuss this with the research team and alternative methods can be arranged.

Additionally, we may be reaching out via text messaging to send you information about your participation in the study. Since texting technology was not built for secure communications, the privacy of information sent through text messaging cannot be guaranteed. Texting should not be your primary source of communication for your healthcare needs. You and your provider can use text messaging as a convenient way to stay in touch. However, we cannot promise that your

text message will be read and responded to within a certain time. By signing below, you accept this risk. Text messaging should never be used for urgent issues or emergencies and you waive any and all claims that may arise against the University of Pennsylvania Health System, employees, contractors, interns, and students resulting from the use or misuse of text messaging. Your provider will make her/his best efforts to use the minimum information necessary in the text messages to reduce the chance of someone else seeing details about you and your health. You can also help reduce privacy risks by putting a screen lock on your device and not sharing your "PIN" or other password to unlock it. I understand that message and data rates may apply. If you wish for us to use a different means to communicate with you during the course of this trial please discuss this with the research team and alternative methods can be arranged.

If you are unable to complete the enrollment process during a virtual or in-person encounter with a study team member, due to not having your social media login credentials at hand or encountering an interruption during your enrollment, the study team will send you up to 4 email reminders in month (using a pennmedicine.upenn.edu email address), with additional phone calls and text messages, if necessary, to share your digital data and social media information and/or complete the baseline survey remotely. To assist you in donating your digital data remotely, we will call or videoconference through Doximity, a HIPAA compliant platform to provide personalized instructions or to troubleshoot any concern or issue.

What are the risks?
The study has minimal risks. Possible risks include the disclosure of potentially sensitive information. Another potential risk in permitting your data to be stored in our research database is the potential risk of loss of confidentiality.

How will I benefit from the study?
There is no benefit to you. However, your participation could help us understand coronary heart disease (CHD) risk factors, which can benefit you indirectly. In the future, this may help other people to prevent CHD.

Will I receive the results of research testing?
Most tests done in research studies are only for research and have no clear meaning for participants. Research results will not be returned to you because the findings are exploratory and are not intended to change the standard of care.

What other choices do I have?
Your alternative to being in the study is to not be in the study.

What happens if I do not choose to join the research study?
You may choose to join the study or you may choose not to join the study. Your participation is voluntary.

There is no penalty if you choose not to join the research study. You will lose no benefits or advantages that are now coming to you, or would come to you in the future. Your primary care doctor will not be upset with your decision.

If you are currently receiving services and you choose not to volunteer in the research study, your services will continue.

When is the study over? Can I leave the study before it ends?
The study is expected to end after all participants have completed all visits and all the information has been collected. The study may be stopped without your consent for the following reasons:

- The PI feels it is best for your safety and/or health-you will be informed of the reasons why.
- You have not followed the study instructions
- The PI, the sponsor or the Institutional Review Board (IRB) at the University of Pennsylvania can stop the study anytime

You have the right to drop out of the research study at any time during your participation. There is no penalty or loss of benefits to which you are otherwise entitled if you decide to do so. Withdrawal will not interfere with your future care.

You may remove your data by visiting https://donate.centerfordigitalhealth.upenn.edu/participant/login. First you will sign into your account and then select the 'Opt Out' button under Actions in the Manage Donated Data. This action removes your data from our database. However, any information gathered before you un-install the application may still be used for analysis.

If you no longer wish to share your social media or health record information with the study, please contact the research team, at Dydhearthealth@Pennmedicine.upenn.edu to ensure you are removed from the study.

To remove your google takeout file and step tracking data, please email Dydhearthealth@Pennmedicine.upenn.edu to permanently delete your file.

If you experience medical or emotional changes related to your involvement in the study, please contact the investigator, Raina Merchant, MD, 3400 Spruce St, Philadelphia, PA 19104, 215-746-8681, raina.merchant@pennmedicine.upenn.edu.

If you no longer wish to be in the research study, please send an email or written notice to the research team, Dydhearthealth@Pennmedicine.upenn.edu

How will my personal information be protected during the study?
We will do our best to make sure that the personal information obtained during the course of this research study will be kept private. However, we cannot guarantee total privacy. Your personal information may be given out if required by law. If information from this study is published or presented at scientific meetings, your name and other personal information will not be used. The Institutional Review Board (IRB) at the University of Pennsylvania will have access to your records.

An exception to confidentiality is if you report child or elder abuse or neglect, or if you report suicidal or homicidal ideation or intent to the research team. Any information about child or elder abuse or intent to harm yourself or others will be reported to the authorities, as required by law.

What may happen to my information collected on this study?
All identifiable data will be kept in a secure, password protected database behind a firewall.

Future Use of Data
Your information will be de-identified. De-identified means that all identifiers have been removed. The information could be stored and shared for future research in this de-identified fashion. It would not be possible for future researchers to identify you as we would not share any identifiable information about you with future researchers. This can be done without again seeking your consent in the future, as permitted by law. The future use of your information only applies to the information collected on this study.

Electronic Medical Record and Release of Study Related Information

What is an Electronic Medical Record and/or a Clinical Trial Management System?

An Electronic Medical Record (EMR) is an electronic version of the record of your care within a health system. An EMR is simply a computerized version of a paper medical record.

A clinical trial management system (CTMS) is used to register your information as a participant in a study and to allow for your research data to be entered/stored for the purposes of data analysis and any other required activity for the purpose of the conduct of the research.

If you are receiving care or have received care within the University of Pennsylvania Health System (UPHS), Penn Medicine, (outpatient or inpatient) and are participating in a Penn Medicine research study, information related to your participation in the research (i.e. laboratory tests, imaging studies and clinical procedures) may be placed in your existing EMR maintained by Penn Medicine Information related to your participation in clinical research will also be contained in the CTMS.

If you have never received care at Penn Medicine and are participating in a Penn Medicine research study that uses Penn Medicine services, an EMR will be created for you for the purpose of maintaining any information produced from your participation in this research study. The creation of this EMR is a requirement of your participation in this study. In order to create your EMR, the study team will need to obtain basic information about you that would be similar to the information you would provide the first time you visit a hospital or medical facility (i.e. your name, the name of your primary doctor, the type of insurance you have). Information related to your participation in clinical research will also be contained in the CTMS. If you have been a patient at Penn Medicine, information from your research participation will be added to your existing medical record.

What may be placed in the EMR?
Information related to your participation in the study (i.e. research enrollment status and contact outreach) will be placed in this EMR maintained by Penn Medicine.

Once placed in your EMR or in the CTMS, your information may be accessible to appropriate Penn Medicine workforce members that are not part of the research team. Information within your EMR may also be shared with others who are determined by Penn Medicine to be appropriate to have access to your EMR (e.g. Health Insurance Company, disability provider, etc.).

UPHS also participates in automated information sharing through Health Information Exchanges (HIEs). HIEs securely share parts of your electronic health record, including research information, with other healthcare organizations involved in your care. This information is shared to improve the quality, safety and efficiency of your healthcare. To request that your health information not be shared through HIEs, please call 215-662-4484.

Will I, as a subject, have access to research related information within the EMR?

The 21st Century Cures Act requires healthcare institutions to allow patients increased access to their electronic medical record. As part of your participation in this research, you will have access to research related information within your EMR through Penn Medicine's patient portal – called MyPennMedicine (MPM).

Will I receive the results of research testing that may be relevant to my health?
Many of tests done in research studies are only for research and have no clear impact on your healthcare. Research results for this study will not be returned to you because they would not be relevant to your healthcare.

Will I have to pay for anything?
You will not have to pay to participate in this study.

Will I be paid for being in this study?
Patients who consent to complete the survey and share digital data will receive $50 in the form of a Greenphire clincard.

Additionally, if you opt to participate in the optional additional survey part of the study, you will receive an extra $10.

If you opt to participate in the optional focus group part of the study, you will receive an additional $50.

You will be entered into a raffle to win an iPad or equivalent device valued at approximately $500. Raffle winners will be chosen at random every 300 study participants enrolled and notified by email and/or phone

What information about me may be collected, used or shared with others?

Data extracted from Facebook may include bio (about me section) text and image posts, time of posting, pages liked, and image captions. Facebook can change privacy settings for applications at any time. Data from Twitter will include: handle, number of followers/following, tweets, geography, time/date, profile. Data extracted from Reddit will include: Reddit posts, comments to posts, your votes "up" or "down" to posts, and the sub-reddits (forums) that you've joined. Data from Instagram will include image posts, time of posting, and image captions. Data extracted from Google takeout includes the search terms, website visited with the URL, YouTube search history, date, and Google fit that includes activities and daily aggregations. If sharing Apple Health, data extracted will include: any data stored or gathered by the Health app and any associated devices, including any Medical ID data, the native iPhone step counter and distance tracker, any data from an Apple Watch, and any data gathered from any third party devices that are syncing to Health app, like a smart scale or blood pressure monitor. (http://osxdaily.com/2019/05/20/export-health-data-from-iphone/). If sharing Fitbit, data extracted will include: Activity, Exercise , Social, Sleep, Coach, Corporate, Logs, Profile, Direct messages, Female health, Sleep score, Friends, Subscriptions. If sharing Garmin, data extracted will include: steps per month for the past 12 months. If sharing S Health, data extracted will include: Daily step count for 34 days prior to export date (https://medium.com/@dimshik100/how-to-extract-your-personal-samsung-health-data-514bbe2331f7). Data extracted from MapMyFitness will include: Date Submitted, Workout Date, Activity Type, Calories Burned, Distance, Workout Time, Average Pace, Max Pace, Average Speed, Max Speed, Average Heart, Rate, Steps, Source, Link *we do not use this as it requires user login to view link results* Data from the medical record will include but not be limited to: demographic information (e.g. date of birth, insurance type, height, weight), medical diagnoses (e.g. history of hypertension, diabetes, cancer, depression), mode of arrival to the hospital (e.g. ambulance or other), labs (e.g. cholesterol, lipid panel, HbA1c), vital signs, medications, procedures, dates and diagnoses from outpatient visits, emergency room visits, hospitalizations. OF NOTE: If you are not a Penn Medicine patient, we will ask you to self-report, in a secure survey link, some of your medical record data as stated above. Health insurance data extracted will include: type of health insurance, health insurance provider, health insurance policy, health insurance group number, and the policy holders name and date of birth will be collected. OF NOTE: If we aren't able to obtain your health insurance policy number, we will either call you to obtain your policy number or you can fill out a secure survey through REDcap that will allow you to input your policy number.

Additional data extracted from our secure data collection platform: device IP addresses.

Why is my information being used?

Your information is used by the research team to contact you during the study. Your information and results of tests and procedures are used to:

- do the research
- oversee the research
- to see if the research was done right
- to evaluate and manage research functions.

Who may use and share information about me?
The following individuals may use or share your information for this research study:

- The Principal Investigator and the Investigator's study team
- Penn researchers and collaborating researchers at other academic institutions with approval from the study team and approval from the Institutional Review Board (IRB)
- Authorized members of the workforce of the UPHS and the School of Medicine and the University of Pennsylvania support offices who may need to access your information in the performance of their duties (for example for research oversight and monitoring)

Who, outside of the School of Medicine, might receive my information?

- No individual or organization

The Principal Investigator or study staff will inform you if there are any additions to the list above during your active participation in the study. Any additions will be subject to University of Pennsylvania procedures developed to protect your privacy.

Once your personal health information is disclosed to others outside the School of Medicine, it may no longer be covered by federal privacy protection regulations.

How long may the School of Medicine use or disclose my personal health information?
Your authorization for use of your personal health information for this specific study does not expire.

Your information may be held in a research database. However, the School of Medicine may not re-use or re-disclose information collected in this study for a purpose other than this study unless:

- You have given written authorization
- The University of Pennsylvania's Institutional Review Board grants permission
- As permitted by law

Can I change my mind about giving permission for use of my information?
Yes. You may withdraw or take away your permission to use and disclose your health information at any time. You do this by sending written notice to the investigator for the study. If you withdraw your permission, you will not be able to stay in this study.

What if I decide not to give permission to use and give out my health information?
Then you will not be able to be in this research study.

e-copy of the informed consent document you signed virtually

You will be given a copy of this Research Subject HIPAA Authorization describing your confidentiality and privacy rights for this study.

By signing this document you are permitting the School of Medicine to use and disclose personal health information collected about you for research purposes as described above.

Who can I call with questions, complaints or if I'm concerned about my rights as a research subject?
If you have questions, concerns or complaints regarding your participation in this research study or if you have any questions about your rights as a research subject, you should speak with Dr. Raina Merchant, (215) 746-8681. If a member of the research team cannot be reached or you want to talk to someone other than those working on the study, you may contact the IRB at (215) 898 2614.

When you sign this form, you are agreeing to take part in this research study. If you have any questions or there is something you do not understand, please ask. You will receive a copy of this consent document.

Printed Name of Subject	Signature of Subject	Date
Printed Name of Witness	Signature of Witness	Date